確率・統計基礎: 大数の法則と集中不等式

森 立平

集中不等式

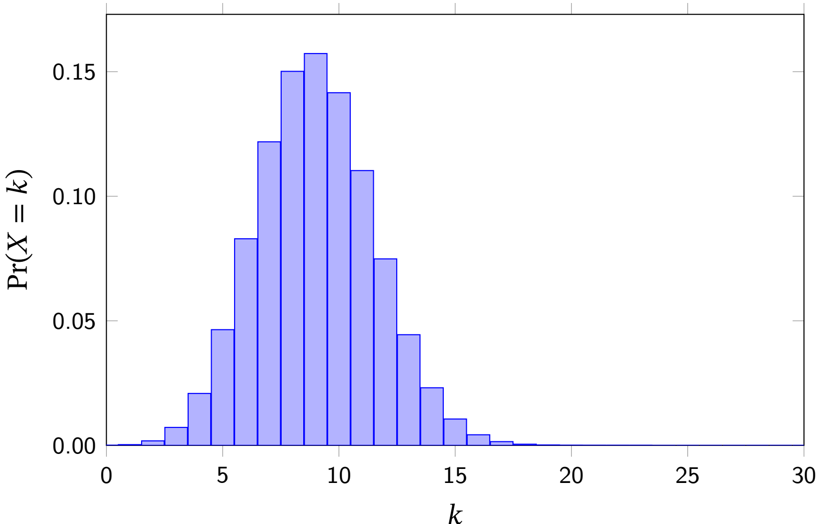
確率変数 $X_1, X_2, \ldots, X_N$ が独立で同一の分布に従うとき、独立同分布 (independently and identically distributed (i.i.d.)) であるという。

このスライドでは $X_1, \ldots, X_N$ は i.i.d. で確率変数 X と同分布であると仮定する。

N 個の i.i.d. 確率変数の平均 $\frac{\sum_{k=1}^{N} X_k}{N}$ が期待値 $\mathbb{E}[X]$ 周辺に集中するというのが大数の法則や集中不等式が示すことである。

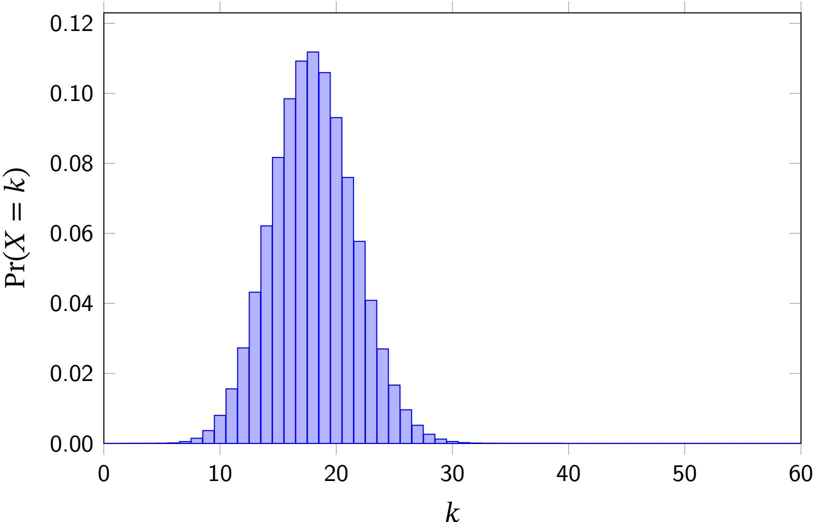
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad {\color{red} n = 30}, \ p = 0.3.$$



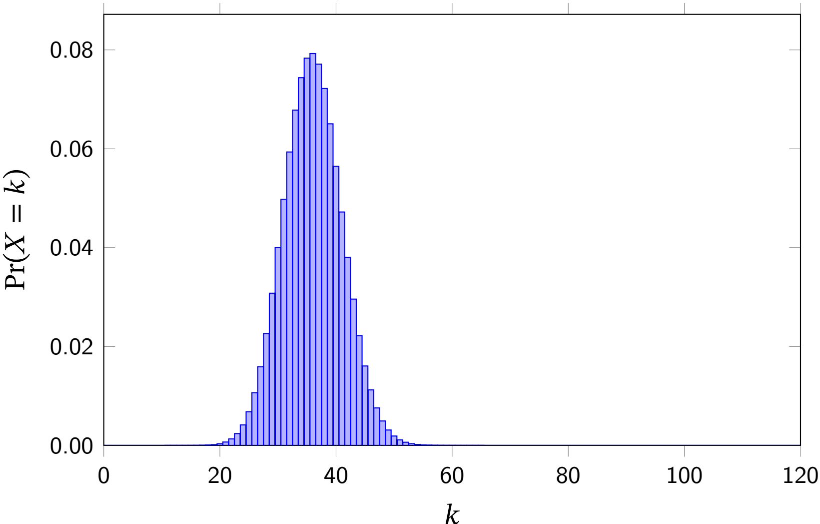
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 60,\ p = 0.3.$$



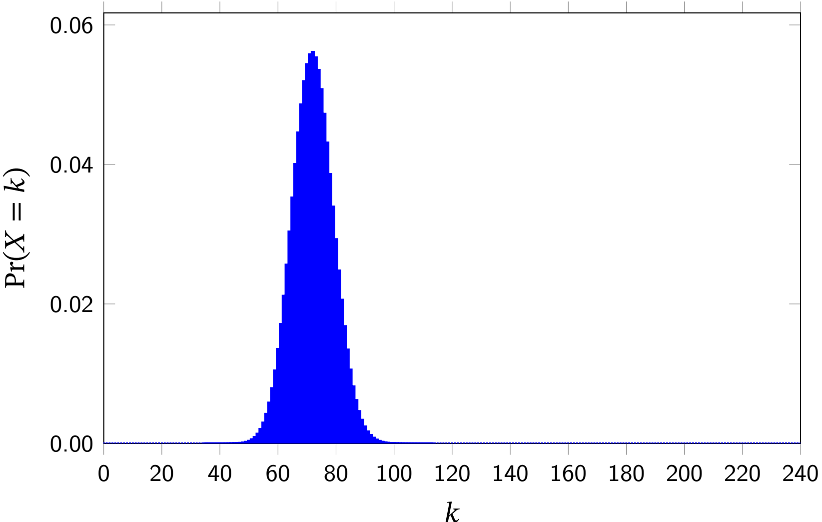
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad {\color{red} n = 120}, \ p = 0.3.$$



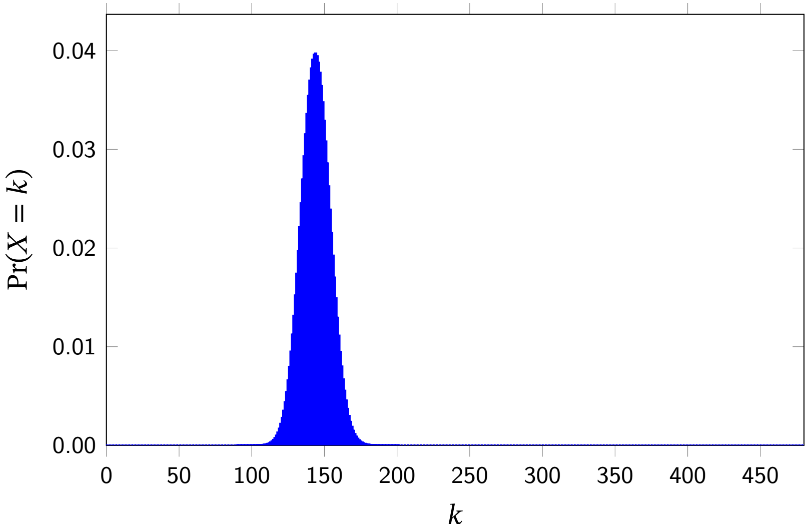
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 240,\ p = 0.3.$$



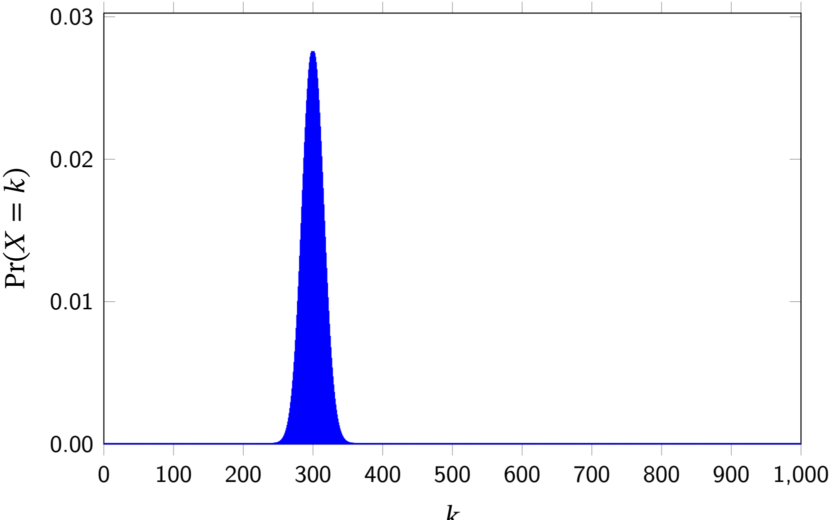
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 480,\ p = 0.3.$$



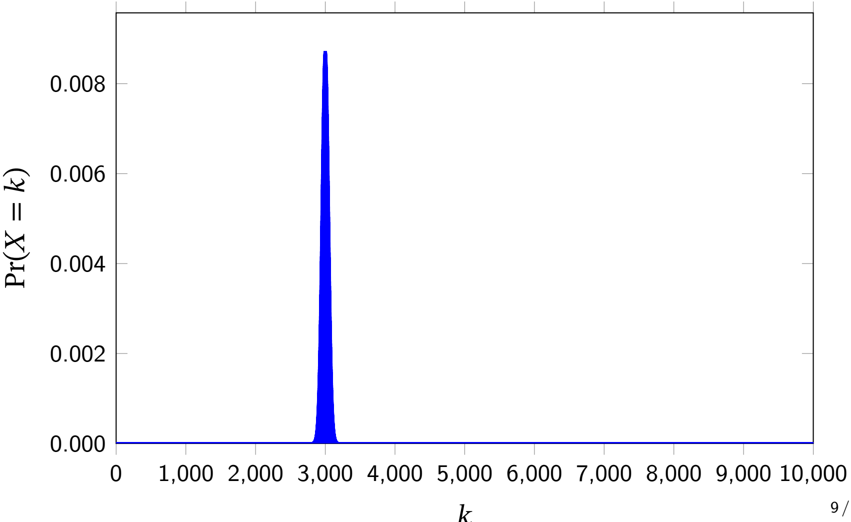
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 1000,\ p = 0.3.$$



二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 10000,\ p = 0.3.$$



集中不等式

Theorem (大数の弱法則)

確率変数 X が分散を持つとする。このとき、任意の実数 $\epsilon > 0$ について

$$\lim_{N\to\infty} \Pr\left(\left|\frac{1}{N}\sum_{k=1}^{N} X_k - \mathbb{E}\left[X\right]\right| \geq \epsilon\right) = 0.$$

Proof.

チェビシェフの不等式より

$$\Pr\left(\left|\frac{1}{N}\sum_{k=1}^{N} X_k - \mathbb{E}\left[X\right]\right| \geq \epsilon\right) \leq \frac{\mathbb{V}\left[\frac{1}{N}\sum_{k=1}^{N} X_k\right]}{\epsilon^2} = \frac{\mathbb{V}\left[X\right]}{\epsilon^2 N} \xrightarrow{N\to\infty} 0.$$

□

チェルノフ上界

Theorem (チェルノフ上界)

任意の $a \in \mathbb{R}$ について

$$\Pr(X \geq a) \leq \frac{M_X(t)}{\mathrm{e}^{at}} = \mathrm{e}^{K_X(t)-at} \qquad \forall t \geq 0$$
$$\Pr(X \leq a) \leq \frac{M_X(t)}{\mathrm{e}^{at}} = \mathrm{e}^{K_X(t)-at} \qquad \forall t \leq 0.$$

Proof.

任意の実数 $t \geq 0$ について

$$\begin{aligned}\Pr(X \geq a) &\leq \Pr\left(\mathrm{e}^{tX} \geq \mathrm{e}^{ta}\right) \\ &\leq \frac{\mathbb{E}\left[\mathrm{e}^{tX}\right]}{\mathrm{e}^{ta}} \qquad \text{(Markov の不等式)} \\ &= \frac{M_X(t)}{\mathrm{e}^{ta}}.\end{aligned}$$

□

任意の実数 $t \leq 0$ について $\Pr(X \leq a) \leq \Pr(\mathrm{e}^{tX} \geq \mathrm{e}^{ta}) \leq \frac{M_X(t)}{\mathrm{e}^{ta}}$.

集中不等式 by チェルノフ上界

Lemma (チェルノフ上界 for 確率変数の和)

任意の $a \in \mathbb{R}$ について

$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \geq a\right) \leq \mathrm{e}^{(K_X(t)-at)N} \qquad \forall t \geq 0$$

$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \leq a\right) \leq \mathrm{e}^{(K_X(t)-at)N} \qquad \forall t \leq 0.$$

Proof.

任意の $t \geq 0$ について

$$\begin{aligned}
\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \geq a\right) &= \Pr\left(\sum_{k=1}^{N} X_k \geq Na\right) \\
&\leq \mathrm{e}^{K_{\sum_k X_k}(t)-Nat} \qquad \text{(チェルノフ上界)} \\
&= \mathrm{e}^{N(K_X(t)-at)}.
\end{aligned}$$

□

チェルノフ上界の最適化

任意の実数 $t \geq 0$ について

$$\Pr(X \geq a) \leq \mathrm{e}^{K_X(t)-at}$$

が成り立つので上界の最小化

$$\inf_{t\geq 0} \mathrm{e}^{K_X(t)-at}$$

を考えたい。log とってから最小化すると

$$\inf_{t\geq 0}\{K_X(t) - at\} = -\sup_{t\geq 0}\{at - K_X(t)\}.$$

キュムラント母関数 $K_X(t)$ のルジャンドル変換

キュムラント母関数の性質

$$K_X(t) = \log M_X(t) = \log \mathbb{E}\left[e^{tX}\right]$$

- $K_X(0) = 0$.
- ある $t > 0$ について $M_X(t)$ が存在する ($\iff$ 有限になる) と仮定すると、$s \in (0, t)$ について

$$\begin{aligned} M_X(s) = \mathbb{E}\left[e^{sX}\right] &= \mathbb{E}\left[e^{sX}\mathbb{1}_{\{X\geq 0\}}\right] + \mathbb{E}\left[e^{sX}\mathbb{1}_{\{X<0\}}\right] \\ &\leq \mathbb{E}\left[e^{tX}\mathbb{1}_{\{X\geq 0\}}\right] + 1 \leq M_X(t) + 1 < \infty \end{aligned}$$

  であるので、$M_X(s)$ も存在する。同様にある $t < 0$ について $M_X(t)$ が存在すると仮定すると $s \in (t, 0)$ について $M_X(s)$ も存在する。よって、M_X や K_X が存在する範囲は 0 を含む区間となる。この区間を $\mathrm{dom}(K_X)$ と表す。
- ある $t_1 < t_2$ について $K_X(t)$ が $t \in (t_1, t_2)$ で存在するとき、この範囲で無限回微分できる (微分と期待値の順番を交換できる)。ルベーグ積分の知識が必要。
- $K_X(t)$ は存在する範囲で凸関数である。

キュムラント母関数の微分と凸性

$$K_X(t) = \log \mathbb{E}\left[e^{tX}\right]$$

$t \in (t_1, t_2)$ でキュムラント母関数が存在すると仮定すると、この範囲で微分と期待値を交換できる。

$$\frac{dK_X(t)}{dt} = \frac{\mathbb{E}\left[Xe^{tX}\right]}{\mathbb{E}\left[e^{tX}\right]}$$

$$\begin{aligned}
\frac{d^2K_X(t)}{dt^2} &= \frac{\mathbb{E}\left[X^2e^{tX}\right]\mathbb{E}\left[e^{tX}\right] - \mathbb{E}\left[Xe^{tX}\right]^2}{\mathbb{E}\left[e^{tX}\right]^2} \\
&= \frac{\mathbb{E}\left[X^2e^{tX}\right]}{\mathbb{E}\left[e^{tX}\right]} - \left(\frac{\mathbb{E}\left[Xe^{tX}\right]}{\mathbb{E}\left[e^{tX}\right]}\right)^2 \\
&= \frac{\mathbb{E}\left[\left(X - \frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}\right)^2 e^{tX}\right]}{\mathbb{E}\left[e^{tX}\right]} \geq 0
\end{aligned}$$

ある確率変数 Z の期待値と分散になっている。また、X が決定的 $(\Pr(X = \mathbb{E}[X]) = 1)$ でない限り、K_X は $\mathrm{dom}(K_X)$ で狭義凸である。

キュムラント

$t \in (-\epsilon, \epsilon)$ でキュムラント母関数 $K_X(t)$ が存在するとき、n 次キュムラントを以下で定義する。

$$\kappa_n := \left. \frac{\mathrm{d}^n K_X(t)}{\mathrm{d}t^n} \right|_{t=0}.$$

- $\kappa_1 = \mathbb{E}[X]$.
- $\kappa_2 = \mathbb{V}[X]$.
- $\kappa_3 = \mathbb{E}\left[(X - \mathbb{E}[X])^3\right]$.
- $\kappa_4 = \mathbb{E}\left[(X - \mathbb{E}[X])^4\right] - 3\kappa_2^2$.
- $\kappa_5 = \mathbb{E}\left[(X - \mathbb{E}[X])^5\right] - 10\kappa_3\kappa_2$.

最初の二つだけ覚えておけば大体 OK。

キュムラント母関数のその他の簡単な性質

任意の独立確率変数 X と Y と $a \in \mathbb{R}$ について

$$K_{X+a}(t) = \log \mathbb{E}\left[\mathrm{e}^{t(X+a)}\right] = \log\left(\mathbb{E}\left[\mathrm{e}^{tX}\right] \cdot \mathrm{e}^{ta}\right) = K_X(t) + at$$

$$K_{aX}(t) = \log \mathbb{E}\left[\mathrm{e}^{t(aX)}\right] = K_X(at)$$

$$K_{X+Y}(t) = \log \mathbb{E}\left[\mathrm{e}^{t(X+Y)}\right] = \log\left(\mathbb{E}\left[\mathrm{e}^{tX}\right]\mathbb{E}\left[\mathrm{e}^{tY}\right]\right) = K_X(t) + K_Y(t)$$

ルジャンドル変換

Definition

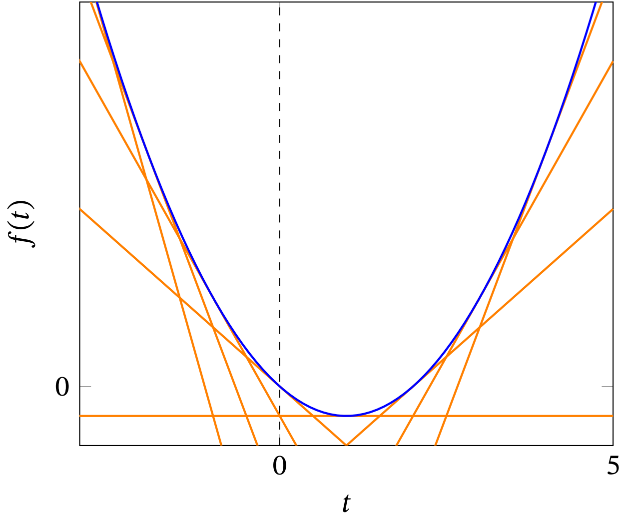
$J \subseteq \mathbb{R}$ を区間とし、関数 $f : J \to \mathbb{R}$ を凸関数とする。f のルジャンドル変換 $f^* : J^* \to \mathbb{R}$ を以下で定義する。

$$f^*(a) = \sup_{t \in J}\{at - f(t)\}$$
$$J^* = \{a \in \mathbb{R} \mid f^*(a) < \infty\}$$

f は凸なので、$at - f(t)$ は t の関数として凹 (上に凸) となる。このとき、

- $f'(t_a) = a$ を満たす $t_a \in J$ が存在するとき、$f^*(a) = at_a - f(t_a)$ である。
- t_a が存在しないときは J の端への極限で sup が達成される。

凸関数と接線

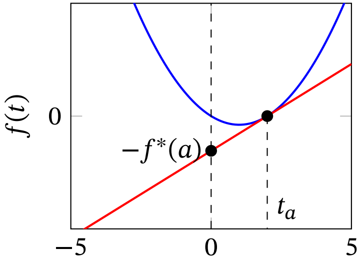


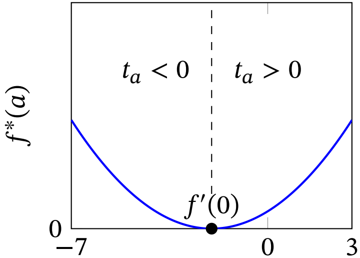
凸関数は接線の集合で表すことができる。接線 $at + b$ は傾き a と切片 b で表せる。凸関数は同じ傾きを持つ接線が一つしかないので、傾きから切片が一意に定まる。この傾きから切片 (の -1 倍) への関数がルジャンドル変換 f^* に他ならない。

ルジャンドル変換

$$f^*(a) = \sup_{t\in J}\{at - f(t)\}$$

$f'(t_a) = a$ を満たす $t_a \in J$ が存在すると仮定する。このとき、
$f^*(a) = at_a - f(t_a)$ である。
f の t_a における接線は $a(t - t_a) + f(t_a)$。この接線の切片は
$-at_a + f(t_a) = -f^*(a)$ である。
例えば $a = f'(0)$ のとき、$t_a = 0$ と取れる。このとき、接線の切片は
$f(0)$ であるので、$f^*(f'(0)) = -f(0)$ であることが分かる。特に
$f(0) = 0$ のとき以下の図のようになる。また、t_a は a について単調非減少であることに注意する。





チェルノフ上界の最適化

Lemma

レート関数 $I_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$ を

$$I_X(a) := \sup_{t \in \mathrm{dom}(K_X)} \{at - K_X(t)\}$$

と定義する。このとき、

$$\sup_{t \geq 0} \{at - K_X(t)\} = \begin{cases} I_X(a) & \textit{if } a > \mathbb{E}[X] \\ 0 & \textit{otherwise.} \end{cases}$$

$$\sup_{t \leq 0} \{at - K_X(t)\} = \begin{cases} I_X(a) & \textit{if } a < \mathbb{E}[X] \\ 0 & \textit{otherwise.} \end{cases}$$

よってチェルノフ上界を最適化することで以下を得る。

$$\Pr(X \geq a) \leq \mathrm{e}^{-I_X(a)} \qquad \forall a > \mathbb{E}[X]$$

$$\Pr(X \leq a) \leq \mathrm{e}^{-I_X(a)} \qquad \forall a < \mathbb{E}[X].$$

最適化されたチェルノフ上界

Theorem (最適化されたチェルノフ上界)

$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \geq a\right) \leq \mathrm{e}^{-I_X(a)N} \qquad \forall a > \mathbb{E}[X]$$
$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \leq a\right) \leq \mathrm{e}^{-I_X(a)N} \qquad \forall a < \mathbb{E}[X].$$

指数関数的な上界が得られた。

例: ベルヌーイ分布

$$\Pr(X=0) = 1-p, \qquad \Pr(X=1) = p$$

$$K_X(t) = \log(pe^t + (1-p))$$

$$I_X(a) = \sup_{t\in\mathbb{R}} \left\{at - \log(pe^t + (1-p))\right\}$$

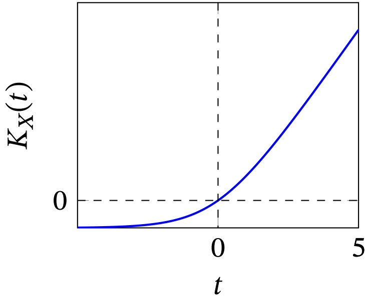
$$a - \frac{pe^{t_a}}{pe^{t_a} + 1 - p} = 0 \iff e^{t_a} = \frac{a(1-p)}{p(1-a)}$$

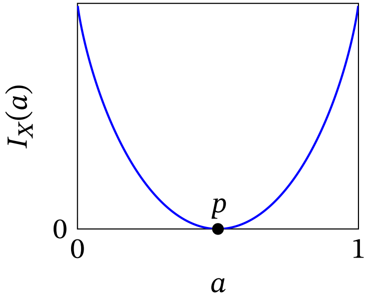
$$\begin{aligned} I_X(a) &= a\log\frac{a(1-p)}{p(1-a)} - \log\left(\frac{a(1-p)}{1-a} + (1-p)\right) \\ &= a\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p} \qquad \text{for } a \in (0,1) \end{aligned}$$

$I_X(0) = -\log(1-p)$, $I_X(1) = -\log p$.

キュムラント母関数とレート関数

$p = 1/2$ の場合





$$K_X(t) = \log(p\mathrm{e}^t + (1-p))$$

$$I_X(a) = a \log \frac{a}{p} + (1-a) \log \frac{1-a}{1-p} \qquad \text{for } a \in [0,1]$$

$0 \log 0$ は 0 とみなす。

例: 正規分布

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for} \quad x \in \mathbb{R}.$$

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{e}^{tx} \, \mathrm{d}x \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}} \mathrm{e}^{\frac{(\mu+\sigma^2 t)^2-\mu^2}{2\sigma^2}} \, \mathrm{d}x \\ &= \mathrm{e}^{\frac{(\mu+\sigma^2 t)^2-\mu^2}{2\sigma^2}} = \mathrm{e}^{\mu t + \frac{\sigma^2 t^2}{2}}. \end{aligned}$$

$$K_X(t) = \mu t + \frac{\sigma^2 t^2}{2}.$$

例: 正規分布

$$
\begin{aligned}
K_X(t) &= \mu t + \frac{\sigma^2 t^2}{2}. \\
I_X(a) &= \sup_{t\in\mathbb{R}} \left\{ at - \left(\mu t + \frac{\sigma^2 t^2}{2} \right) \right\} \\
&= \sup_{t\in\mathbb{R}} \left\{ -\frac{\sigma^2}{2} t \left(t - \frac{2(a-\mu)}{\sigma^2} \right) \right\} \\
&= \frac{(a-\mu)^2}{2\sigma^2}.
\end{aligned}
$$

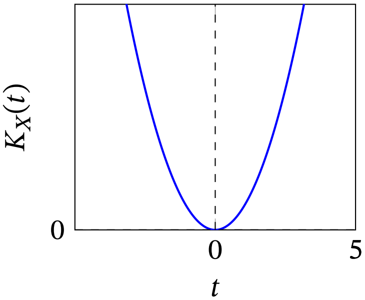
任意の $\epsilon > 0$ について

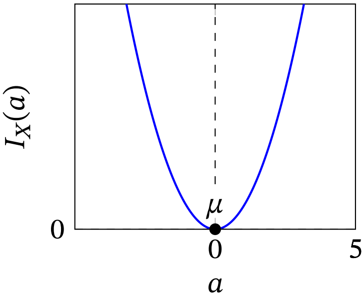
$$
\Pr\left(\frac{1}{N} \sum_{k=1}^{N} X_k \geq \mu + \epsilon \right) \leq \mathrm{e}^{-\frac{\epsilon^2}{2\sigma^2} N}
$$

$$
\Pr\left(\frac{1}{N} \sum_{k=1}^{N} X_k \leq \mu - \epsilon \right) \leq \mathrm{e}^{-\frac{\epsilon^2}{2\sigma^2} N}.
$$

キュムラント母関数とレート関数

$\mu = 0,\ \sigma^2 = 1$ の場合





$$K_X(t) = \mu t + \frac{\sigma^2 t^2}{2}$$

$$I_X(a) = \frac{(a-\mu)^2}{2\sigma^2}$$

例: ポアソン分布

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \text{for} \quad k \in \mathbb{Z}_{\geq 0}.$$

$$\begin{aligned} M_X(t) &= \sum_{k \geq 0} \frac{\lambda^k e^{-\lambda}}{k!} e^{tk} \\ &= e^{\lambda e^t - \lambda} \\ K_X(t) &= \lambda(e^t - 1) \end{aligned}$$

例: ポアソン分布

$$K_X(t) = \lambda(\mathrm{e}^t - 1)$$

$$I_X(a) = \sup_{t\in\mathbb{R}} \left\{at - \lambda(\mathrm{e}^t - 1)\right\}$$

$$a - \lambda \mathrm{e}^{t_a} = 0 \iff t_a = \log\frac{a}{\lambda}$$

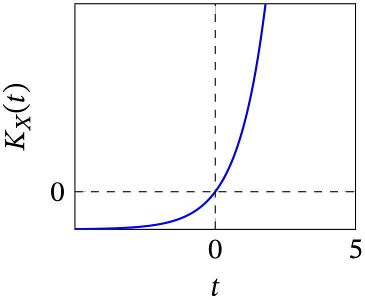
$$\begin{aligned} I_X(a) &= \sup_{t\in\mathbb{R}} \left\{at - \lambda(\mathrm{e}^t - 1)\right\} \\ &= a\log\frac{a}{\lambda} - a + \lambda \qquad \text{for } a \geq 0. \end{aligned}$$

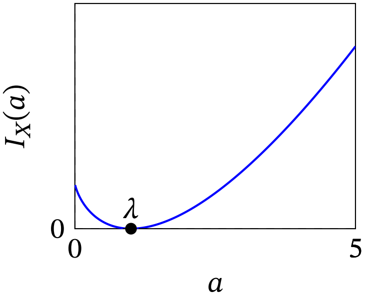
$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \geq c\lambda\right) \leq \mathrm{e}^{-\lambda(c\log c - c + 1)N} \qquad \forall c > 1$$

$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \leq c\lambda\right) \leq \mathrm{e}^{-\lambda(c\log c - c + 1)N} \qquad \forall c < 1$$

キュムラント母関数とレート関数

$\lambda = 1$ の場合





$$K_X(t) = \lambda(\mathrm{e}^t - 1)$$

$$I_X(a) = a \log \frac{a}{\lambda} - a + \lambda \qquad \text{for } a \geq 0.$$

例：指数分布

$$p(x) = \lambda \mathrm{e}^{-\lambda x} \qquad \text{for} \quad x \geq 0.$$

$$\begin{aligned} M_X(t) &= \int_0^\infty \lambda \mathrm{e}^{-\lambda x} \mathrm{e}^{tx} \, \mathrm{d}x \\ &= \left[\frac{\lambda}{t-\lambda} \mathrm{e}^{(t-\lambda)x} \right]_0^\infty \\ &= \begin{cases} \frac{\lambda}{\lambda - t} & t < \lambda \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

$$K_X(t) = \begin{cases} \log \frac{\lambda}{\lambda - t} & t < \lambda \\ +\infty & \text{otherwise.} \end{cases}$$

例: 指数分布

$$K_X(t) = \log \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda.$$

$$I_X(a) = \sup_{t<\lambda} \left\{ at - \log \frac{\lambda}{\lambda - t} \right\}$$

$$a - \frac{1}{\lambda - t_a} = 0 \iff t_a = \lambda - \frac{1}{a}$$

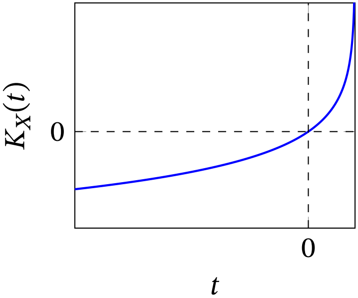
$$\begin{aligned} I_X(a) &= \sup_{t<\lambda} \left\{ at - \log \frac{\lambda}{\lambda - t} \right\} \\ &= a\lambda - 1 - \log(a\lambda) \qquad \text{for } a > 0 \end{aligned}$$

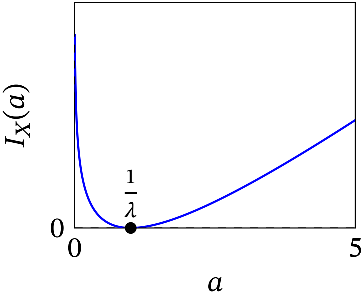
$$\Pr\left(\frac{1}{N} \sum_{k=1}^{N} X_k \geq \frac{c}{\lambda} \right) \leq \mathrm{e}^{-(c-1-\log c)N} \qquad \forall c > 1$$

$$\Pr\left(\frac{1}{N} \sum_{k=1}^{N} X_k \geq \frac{c}{\lambda} \right) \geq \mathrm{e}^{-(c-1-\log c)N} \qquad \forall c < 1.$$

キュムラント母関数とレート関数

$\lambda = 1$ の場合





$$K_X(t) = \log \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda$$

$$I_X(a) = a\lambda - 1 - \log(a\lambda) \qquad \text{for } a > 0$$

大偏差原理 Lv1

レート関数 $I_X(a)$ は**最適な指数**であり、これ以上改善することはできない。証明は難しいので紹介しない。

Theorem (クラメールの定理)

$$\lim_{N\to\infty}\frac{1}{N}\log\Pr\left(\frac{1}{N}\sum_{k=1}^{N}X_k \geq a\right) = -I_X(a) \qquad \forall a > \mathbb{E}[X]$$

$$\lim_{N\to\infty}\frac{1}{N}\log\Pr\left(\frac{1}{N}\sum_{k=1}^{N}X_k \leq a\right) = -I_X(a) \qquad \forall a < \mathbb{E}[X]$$

サブガウシアン

キュムラント母関数やレート関数が綺麗に計算できればよいのだが、必ずしもそうとは限らない。

Definition

X が代理分散 $s^2 > 0$ についてサブガウシアン $\overset{\text{def}}{\iff}$ 任意の $t \in \mathbb{R}$ について

$$K_X(t) \leq \mathbb{E}[X]\, t + \frac{s^2 t^2}{2}.$$

キュムラント母関数を正規分布のキュムラント母関数で上から抑えることで、正規分布の場合と同じ形の不等式が得られる。
X が代理分散 $s^2 > 0$ についてサブガウシアンのとき、任意の $\epsilon > 0$ について

$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \geq \mathbb{E}[X] + \epsilon\right) \leq \mathrm{e}^{-\frac{\epsilon^2}{2s^2}N}, \quad \Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \leq \mathbb{E}[X] - \epsilon\right) \leq \mathrm{e}^{-\frac{\epsilon^2}{2s^2}N}.$$

有界な確率変数

Lemma (Hoeffding's lemma)

確率変数 X が

$$\Pr(X \in [a,b]) = 1$$

を満たすとする。このとき X は代理分散 $\frac{(b-a)^2}{4}$ についてサブガウシアンである。つまり、

$$K_X(t) = \log \mathbb{E}\left[e^{tX}\right] \le \mathbb{E}[X]\, t + \frac{(b-a)^2 t^2}{8}.$$

Proof.

X が有界であることから、$K_X(t)$ は任意の $t \in \mathbb{R}$ について存在する。テイラーの定理よりある $\theta \in (0,t)$ が存在して、($t < 0$ のときは $\theta \in (t,0)$)

$$K_X(t) = \mathbb{E}[X]\, t + K_X''(\theta)\frac{t^2}{2}$$

ここで、$K_X''(\theta)$ はある確率変数 $Z \in [a,b]$ の分散 $\mathbb{V}[Z]$ である。 □

Hoeffding's lemma の証明

証明の続き.

$$K_X''(t) = \mathbb{E}\left[\left(X - \frac{\mathbb{E}\left[X e^{tX}\right]}{\mathbb{E}\left[e^{tX}\right]}\right)^2 e^{tX}\right] / \mathbb{E}\left[e^{tX}\right].$$

具体的には任意の $f\colon \mathbb{R} \to \mathbb{R}$ について

$$\mathbb{E}\left[f(Z)\right] = \frac{\mathbb{E}\left[f(X)e^{\theta X}\right]}{\mathbb{E}\left[e^{\theta X}\right]}$$

という確率変数 Z について、$K_X''(\theta) = \mathbb{V}\left[Z\right]$ である。
例えば X の確率密度関数が $p(x)$ のとき、Z の確率密度関数は
$p(x)\frac{e^{\theta x}}{\int_{-\infty}^{\infty} p(x)e^{\theta x}\,dx}$ である。$\Pr(Z \in [a, b]) = 1$ である。

$$\mathbb{E}\left[(Z - x)^2\right] = x^2 - 2\mathbb{E}\left[Z\right]x + \mathbb{E}\left[Z^2\right] \geq \mathbb{V}\left[Z\right] \qquad \forall x \in \mathbb{R}$$

より、$x = \frac{a+b}{2}$ を代入すると、$\mathbb{V}\left[Z\right] \leq \mathbb{E}\left[\left(Z - \frac{a+b}{2}\right)^2\right] \leq \left(\frac{b-a}{2}\right)^2$. □

Hoeffding 不等式

Theorem (Hoeffding 不等式)

確率変数 X が

$$\Pr(X \in [a, b]) = 1$$

を満たすとする。任意の $\epsilon > 0$ について

$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \geq \mathbb{E}\left[X\right] + \epsilon\right) \leq \mathrm{e}^{-\frac{2\epsilon^2}{(b-a)^2}N}$$

$$\Pr\left(\frac{1}{N}\sum_{k=1}^{N} X_k \leq \mathbb{E}\left[X\right] - \epsilon\right) \leq \mathrm{e}^{-\frac{2\epsilon^2}{(b-a)^2}N}.$$

おまけ: テイラーの定理

Theorem (テイラーの定理)

実関数 $f : \mathbb{R} \to \mathbb{R}$ が $a \in \mathbb{R}$ で $k \geq 1$ 回微分可能なとき、ある関数 $h_k : \mathbb{R} \to \mathbb{R}$ で $\lim_{x \to a} h_k(x) = 0$ であるものが存在し、

$$f(x) = \sum_{s=0}^{k} \frac{f^{(s)}(a)}{s!}(x-a)^s + h_k(x)(x-a)^k.$$

また、$f : \mathbb{R} \to \mathbb{R}$ が $[a, x]$ 上 で $k+1$ 回微分可能であるとすると、ある $c \in (a, x)$ が存在し、

$$f(x) = \sum_{s=0}^{k} \frac{f^{(s)}(a)}{s!}(x-a)^s + \frac{f^{(k+1)}(c)}{(k+1)!}(x-a)^{k+1}.$$

課題

1. 表が出る確率が 1/2 のコインを N 回投げて $0.6N$ 回以上表が出る確率を e^{-cN} という形で上から抑えるときの最大の $c > 0$ をもとめよ。

2. $f(t) = t \log t - t + 1 \quad \text{for } t > 0$ のルジャンドル変換 $f^*(a)$ をもとめよ。

3. $n > 1$ について $f(t) = \frac{1}{n}|t|^n \quad \forall t \in \mathbb{R}$ のルジャンドル変換 $f^*(a)$ をもとめよ。